



Munich Personal RePEc Archive

Open Data: Policy and Implementation in Bulgaria

Gerunov, Anton

Sofia University "St. Kliment Ohridski"

2015

Online at <https://mpra.ub.uni-muenchen.de/68799/>

MPRA Paper No. 68799, posted 13 Jan 2016 14:21 UTC

Open Data: Policy and Implementation in Bulgaria

Anton Gerunov

Council of Ministers of the Republic of Bulgaria and Sofia University “St. Kliment Ohridski”

125 Tsarigradsko Shosse Blvd, 1113 Sofia, Bulgaria

gerunov@uni-sofia.bg

Abstract: This paper provides an overview of Bulgaria’s policy on open data by using the framework of a novel Open Data stage model. We also present a detailed overview the data landscape of the public sector in the country and outline the first milestones in its efforts to publish some of that data for reuse. The paper outlines the limited knowledge about true open data impact and proposes a few approaches to measure it formally in order to critically assess the usefulness of opening data.

Keywords: Open data, OGD, data audit, open data portal

1 Introduction

Open Data has received enormous attention as an integral part of the Open Government Movement. The concept revolves around the need of public bodies to share the information they collect in the regular course of their activities with the public so that this data can be reused and unlock additional economic and social value [1]. Open data encompasses much more than open government data (OGD) alone but the public sector seems to be the frontrunner in opening its data at this stage, followed closely by the scientific community in opening research data and promoting open science [2]. The private and the non-profit sector will likely be more proactive over the next years.

The main driving forces behind opening data concentrate around three major benefits that it is supposed to bring to the economy and the public space. These are the following [1], [3], [4]:

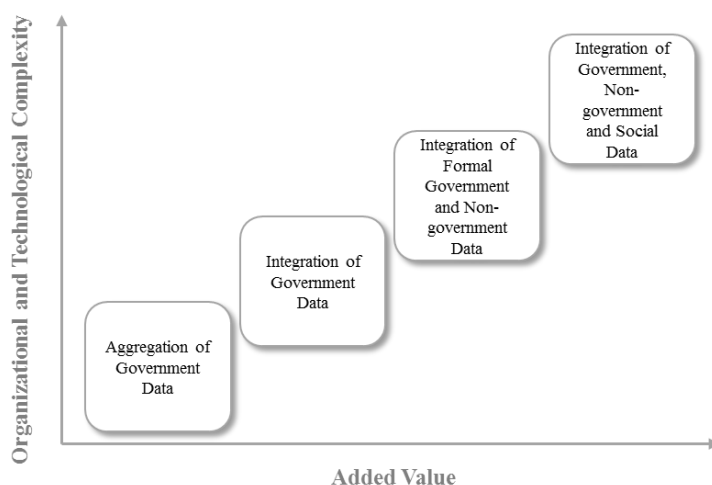
- Unlocking economic value by stimulating private sector innovation, providing for better business decision-making, and promoting the data science industry.
- Providing political and social benefits by increasing the public’s capacity to monitor government and participate more effectively in the formulation of public policy.
- Realizing operation improvement in the public sector by removing duplications, improving the access to information and optimizing processes through benchmarking and reengineering.

Those potential benefits have led many countries to devise and implement open data policies in the hope of unlocking its large potential, and Bulgaria stands to be one of those countries. This paper will review the country's efforts within the framework of an Open Data stage model in Section 2 and report on the data landscape of the public sector in Section 3. It will also explore the impact of open data initiatives in Section 4, and give policy recommendations in Section 5. Section 6 concludes.

2 A Stage Model of Open Data

The theory and practice of OGD is relatively new and therefore few theoretical models are available to put a conceptual framework around a country's efforts. One can borrow from neighboring fields such as electronic government [5] but these are not fully applicable. For the purposes of this paper we are going to use a specific Open Data Stage model, developed by Kalampokis et al. [6]. This model is based on four main stages of opening data, each with increased sophistication and difficulty, but also with increased potential of unlocking value. The model is presented graphically in Figure 1.

Figure 1: A Stage Model of Open Data, Kalampokis et al.



The first stage consists of aggregating government data, whereby public sector agencies and units create inventories of their data, export it, and publish it. The format used is usually a well-known machine-readable format such as csv, json, xml, and others. The first stage is crucial as it has to overcome a number of technological, legislative, economic, and cultural barriers as many organizations are both unable and sometimes unwilling to freely share their data. To achieve success during this stage it is crucial to modify the data sharing culture and

ensure sustainability for the overall process. There is some economic value to be unlocked by this stage as the public and private companies have access to previously unutilized data.

The second stage takes the data sharing efforts a step further by providing for integration of OGD contained in different databases. Thus it turns information into knowledge by creating a unified data view focused on a given object (company, region, government entity, etc.) by connecting all data in the public sector that pertains to it. This stage faces significant technological and organizational barriers. On the technological side, data formats must be compatible with each other, so that integration is possible. Currently, the Linked Open Data paradigm seems to be a leading contender to achieve this. On the organizational side challenges are even bigger. Data integration will likely show duplicated data and will reveal errors and inconsistencies in the information the public sector uses for decision-making. This may lead to erosion of public trust and inspire increased corrective action – two things public sector bodies would rather avoid. We should note that initially integration is likely to be only partial and only as time passes it will be increasingly complete.

The third stage seeks to create even further opportunities by integrating formal government data with formal non-government data coming from the private sector, the media, or civil society organizations. A lot of non-public entities collect and maintain databases that can be potentially useful as they provide further knowledge for a given object of interest. Integrating OGD with those provides for even greater economic value. This stage is particularly challenging as it is the first that requires concerted efforts outside the public sector and thus outside the reach of government executive authority. Private companies and NGOs need to be convinced to share their data and it must also be processed so that integration with existing OGD is possible. Further, this pursuit may not be fruitful for all sorts of data since the use cases are not always clear and even in the event of positive economic benefit the implementation costs may be prohibitively high.

The fourth and final stage provides for integration of formal government and non-government data with social data, including information from social networks. Social data is created and voluntarily shared by citizens and often expresses opinion, belief, attitudes, and values. A large quantity of this data is streamed live through networks such as Facebook and Twitter and is sometimes accessible through their APIs. Such data integration can make for very sophisticated queries and can serve to spur innovation in economic production and government. This, however, comes at the prices of very high implementation costs and raises questions about privacy and control. The latter two stages – three and four – at this point seem

difficult to achieve at a large scale but some instances of individuals and companies combining data from different sources abound.

Using the Stage Model as a framework, one can classify Bulgaria as belonging to the first stage, making initial steps with data inventories and beginning to publish key datasets of public interest on its Open Data Portal. At this initial stage the country faces its associated challenges – lack of data overview, difficulty in standardizing and publishing data, reluctance of public bodies to share information, and still limited use cases. In the next section we present the results of a data audit in the country to outline the starting position of OGD initiatives and then outline the concrete steps Bulgaria has taken to open data and unlock their value.

3 Data Resources in Bulgaria

Knowledge of the data landscape in the public sector is crucial for an open data initiative to succeed. This is sometimes challenging, as the government is not a monolithic entity but rather a collection of functionally organized administrations with set goals and related data they collect for their purposes. There are 576 administrative units that offer services to citizens, as registered in the Administrative Registry. Additionally, there are more than 2,363 different administrative services provided by these units, and most of them have informational requirements [7]. Data collection and access is either mandated by law, dictated by practical necessity, or done for historical reasons. This section outlines the key results from a full information audit of the public administration in Bulgaria and shows how it informed OGD policy and implementation.

3.1 Key Results from the Data Audit in Bulgaria

The preliminary data audit was done over the last quarter of 2014 and the first quarter of 2015 by requesting a complete data questionnaire from every single administrative unit. Since this was an initiative led by the Council of Ministers it had a very high response rate well over 90%, or a total of 564 responses received. In addition to that further research was undertaken to add more informational sources mandated by in order to complete the list. The audit shows a total of 8,156 different data sources in the administrations that are kept in over 1,300 server spaces and numerous work stations. The primary way of storing public sector data is via means of an internal server with 31% of respondents mentioning this. This is closely followed by storing data on external drives, paper and other means (30%), and on local workstations (25%). Outsourcing data storage remains unpopular with only 10% of administrative units using external servers or hosting. This indicates a culture of reluctance to disclose data

and preference for internal handling and may be interpreted as symptomatic of a reserved attitude towards opening data.

Figure 2: Storage of Data in the Public Sector in Bulgaria

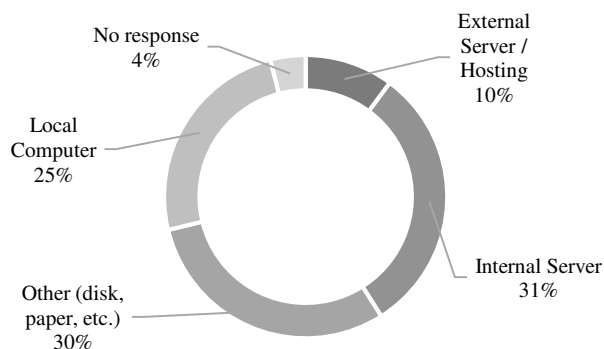


Table 1 presents an overview of the formats, used by administrations. The first striking conclusion is that an overwhelming amount of data is stored in formats that are difficult to process for further analysis. Word files amount to 9.1% of all data, pdf files – for another 4.1%, and fully 29.3% of all data is still stored exclusively on paper. This includes not only internal and external documents but also registries that are used in the process of service delivery. Structured formats also feature prominently, with almost 22% of information stored in Excel files. Fully machine readable files, that are ready for further processing include .mdb files (4.4%), .html files (4.1%), .xml and other databases (1.5%). About 19% of administrations could not exactly specify their data storage and opted for the answer “Other”.

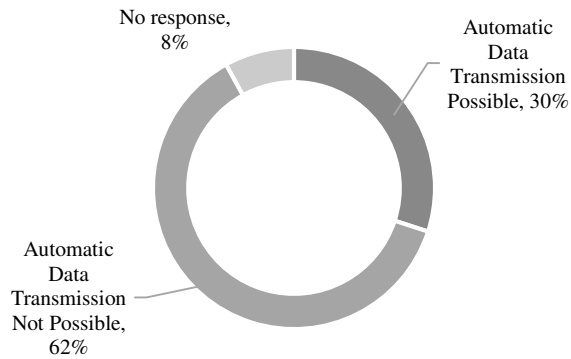
Table 1: Data formats used in the public sector

.doc/ .docx	.html	.mdb	.pdf	.xls/ .xlsx	.xml	Data Base	Paper	Other	No reponse	Total
741	334	362	338	1786	76	49	2396	1589	484	8171
9.1%	4.1%	4.4%	4.1%	21.9%	0.9%	0.6%	29.3%	19.6%	5.9%	100.0%

It is of particular note that information is still stored predominantly on paper, or in text formats. Even formats that should be machine-readable or easily export to a machine-readable form are often structured in such a way to hinder the process. This holds particularly true for Excel files, which are sometimes formatted in view more of their visual appeal and less in terms of interoperability (including merged cells, inserted columns and rows for formatting purpose, using visual templates, etc.) In total, this makes for a large number of files that are not truly machine-readable – 46% of registries are not, while only 41% of them are. For the rest, no response was given. Automatic data access is crucial for the provision of services and

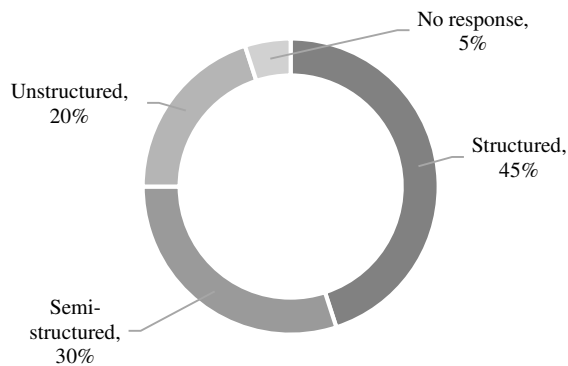
for the improvement of governance processes. It is also no less important for the OGD initiative as it makes for easier extraction and updates. Most of the data in the public sector, however, cannot be accessed automatically (Fig. 3).

Figure 3: Automatic Data Transmission



The overall conclusion is that less than half of the data audited consists of structured data – be it textual or numeric. There is a large chunk of semi-structured data, and about one fifth is completely unstructured (Fig. 3). This calls for significant efforts at structuring and standardizing data before opening and sharing it with the public. The process is likely to be hindered by the wide unavailability of meta-data – fully 64% of the respondents say that no metadata is available, and only 28% of public sector information is accompanied by some metadata.

Figure 4: Structure of the data in the public administration

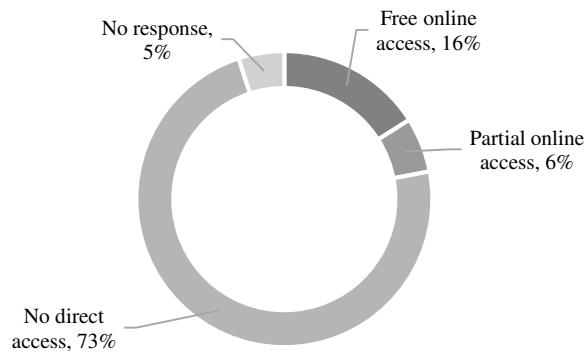


A final question of key interest for the OGD is whether any of this data is accessible to the public. If data is already published, its aggregation on a single access point brings the benefit of convenience and increases awareness. If the data has never been publicly available,

its publication has the potential to unlock both social and economic benefits as actors in the public domain experience a tangible increase in disposable information.

A small fraction of public sector information is available to the public – only 16% of data has free online access. An additional 6% of data is published only in part at the discretion of the administration. Fully 73% of data sources remain unavailable for reuse by private agents and this comprises a measure of OGD’s potential to generate business and social innovation. The large share of undisclosed information clearly outlines the need for a targeted and comprehensive open data policy that will ensure effective reuse of data and provide for sustainability of this process.

Figure 5: Data availability for public access



3.2 Milestones in Opening Data

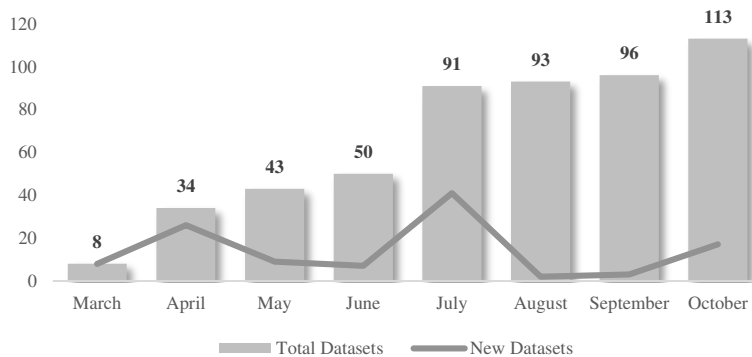
Bulgaria has become active in the process of opening public sector data relatively late in respect to other EU countries. In terms of legislation, the grounds for publishing OGD at the EU level is given in Directive 2013/37/EC, which is transposed in Bulgarian legislation by means of modification of the Freedom to Information Act. A demonstration portal for open data is active since October 2014 and policy efforts to populate it have been undertaken since then. National OGD policy has been enshrined as country reform plans in pursuit of Bulgaria’s second Open Government Action plan under the international initiative Open Government Partnership (OGP). In it the country pledged to publish extensive public sector information. Initially the list of data to be published was prioritized by an Open Data task force in the Council of Ministers.

The task force gave particular weight to the relative importance of data, its ability to disclose government expenditure, to inform policy debate on issues such as healthcare, education, environment, law enforcement, and to aid in improving public sector efficiency. Thus the

task force came up with a list of 119 priority datasets to be opened that included the public procurement register and budget payments, educational performance statistics, air pollution data, vacant jobs data, road accidents and crime statistics, business and demographic statistics, donors in kind to political parties, etc. The list was adopted by the Council of Ministers with CM Act 103/2015 and public sector organization were required to open those datasets and publish them on the single point of access – the online portal *opendata.government.bg*.

Following the Council for Administrative Reform schedule, public sector units started uploading data since March 2015. Progress has been significant (Fig. 6) and the total number of datasets by the end of 2015 will probably go beyond the initially planned 119. Further prioritization is also possible through the feedback form on the portal that allows users to request data of interest to them.

Figure 6: Number of available datasets on the open data portal, Q1-Q3.2015



In conjunction with the top down approach, the government has mounted initiatives to popularize the use and reuse of OGD. Among those are press conferences and seminars such as “A Date with Data” in July 2015, and a number of trainings for about 100 data administrators.

4 Open Data Impact

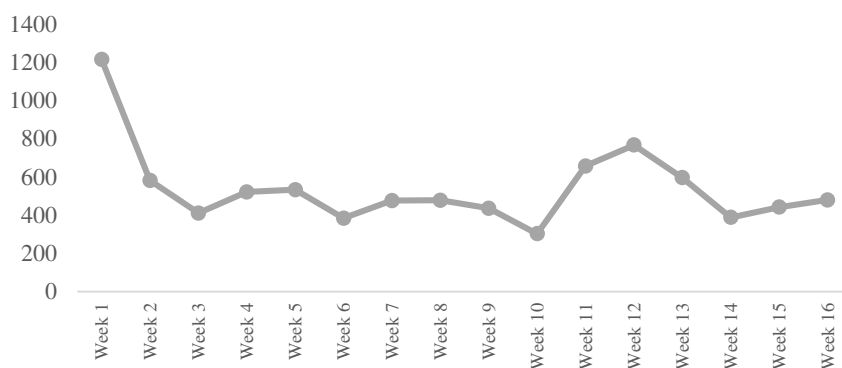
The impact of opening up data is often debated and espoused as the primary reason for publishing OGD [1], [3]. While recourse to its economic and democratic impact is seen as useful driver for publicizing more data it is rarely easy to quantify the impact this initiative has on business and society. So far, efforts at measuring impact has been mixed and unable to produce concrete results on the usefulness of OGD [8]. The crux of the issues lies in the fact that merely opening up datasets does not automatically mean that the public can use them meaningfully or that business can profitably utilize them. Apart from access, the impact of open data depends crucially on engagement, ability to analyze and draw conclusions from

information, and a suitable institutional and economic environment that is receptive of such innovation. In fact, barriers to usage of open data are sometimes seen as so high that some authors argue that OGD empowers the already empowered – the highly educated persons and sophisticated businesses that can extract value from public information [9]. All this is likely to put real world open data impact as smaller and more unequal than usually discussed in public policy circles.

4.1 Usage Metrics and Anecdotal Evidence

A common approach for measuring policy success is through the sheer number of available datasets as this is supposed to capture the availability of public data for reuse. As of the beginning of November 2015, Bulgaria has published a total of 113 datasets on its single point of access and continues towards the target of 119. Such a measure misrepresents the impact of data as it fails to discern both the public’s exposure to information, as well as its quality and usefulness. How often users leverage OGD can easily be measured by a metric of the open data portal’s popularity such as visits, unique users, or average session duration. By this measure open datasets in Bulgaria score relatively well with an average monthly visits of 2,167 (Fig. 7), but still some improvement is possible. For a period of 4 months from 15 July 2015 to 31 October, the site has had 8,668 sessions and 6,188 users. Of those 72.3% were new visitors, while the rest were return visitors.

Figure 7: Active Session on Open Data Portal, 15.07.2015-31.10.2015



The quality, usefulness and impact of information is harder to gauge. While there is some anecdotal evidence of effective data usage, total impact remains elusive. Despite this, a few success stories have proliferated giving examples of OGD potential. One of them is connected with the release of new voter registration in the run-up to the October 2015 municipal elections which showed a huge abnormal spike in registrations and propelled authorities to

take action against it. While informative such case studies need to be supplemented by a fuller and more comprehensive account of OGD's usefulness.

4.2 True Impact Metrics

Impact metrics need to quantify both economic and political benefits brought about by the totality of open data, and also take account of the distribution of those benefits. We can outline two major approaches to measuring this impact:

- **Macro-level approaches** – since OGD is supposed to stimulate information and improve the public environment, it should be the case that it is associated with a measure of technological development such as Total Factor Productivity (TFP). A possible approach is to use a general linear model with TFP as dependent variable, and a measure of OGD and a vector of controls as independent variables. While imperfect, this approach can give an estimate whether OGD has a transformative power for the overall economy and serve as a useful guide for policy discussion.
- **Meso-level approaches** – opening data in a specific sector should bring notable improvement in it, which can be seen in some pre-determined data indicators. For example, opening procurement data should lead to more transparency and less corruption and thus lower the price for reference orders. Other causal effects can be ruled out or controlled for using the GLM model outlined above. Such metrics have the power to provide the very concrete benefits for opening data and will be especially useful for improving the efficiency and effectiveness of public sector units and organizations.
- **Micro-level approaches** – these focus on a specific datasets or groups of datasets, and follow them through their lifecycle. By doing this, the researcher gets a full and nuanced picture of usage, impact, and benefit distribution. The most common micro-level approach is the case study whereby each OGD dataset usage is described in detail, giving the context and measuring benefits to different stakeholders [10]. Case studies generally use a mixed methods design and serve as an excellent illustration of OGD potential. They can thus be leveraged as a powerful argument in favor of openness. The main issues with this approach is that it fails to scale well and is suffering from observer bias.

The method of choice for measuring impact naturally differs across situations and has to adapt to the context of specific data openness. What is most important is not to overlook this key aspect of OGD policy. It is indeed difficult to manage something that is not measured.

5 Conclusions

The paper makes an overview of Bulgaria's open data policy and puts it in the context of a stage model to outline possible future directions. For a very short period of time the country has disclosed a relatively large number of datasets with large economic and social potential. While data usage and popularization is still in need of improvement, there is already some anecdotal evidence of OGD's impact.

Despite the short period of focused policy, Bulgaria has reached a level of open data maturity which allows for a number of important further developments. First, it needs to spell out a specific Open Data strategy outside the framework of OGP to underline its political commitment. Second, more targeted efforts at measuring OGD impact will serve as important drivers for sustainability and expansion of the policy efforts. Third, and most important, the country needs to make more confident strides to the second stage of OGD development by linking government datasets in order to unlock even more value. In conclusion, if fast-paced progress is maintained, OGD has the potential to serve as a transformative power for the public and private sectors alike.

References

- [1] Janssen, M., Charalabidis, Y., & Zuiderwijk, A. (2012). Benefits, adoption barriers and myths of open data and open government. *Information Systems Management*, 29(4), 258-268.
- [2] Molloy, J. C. (2011). The open knowledge foundation: open data means better science. *PLoS Biol*, 9(12), e1001195.
- [3] Huijboom, N., & Van den Broek, T. (2011). Open data: an international comparison of strategies. *European journal of ePractice*, 12(1), 4-16.
- [4] Davies, T. (2010). *Open data, democracy and public sector reform. A look at open government data use from data.gov.uk*. UK: Open Data Impacts.
- [5] Directorate General for Information Society and Media. (2009). *Smarter, Faster, Better eGovernment. 8th eGovernment Benchmark Measurement*. Brussels: European Commission.
- [6] Kalampokis, E., Tambouris, E., & Tarabanis, K. (2011). Open government data: A stage model. *Electronic government* (pp. 235-246). Springer Berlin Heidelberg.
- [7] Council for Administrative Reform. (2015). *Report on the Public Administration, 2014*. Sofia: Council of Ministers.
- [8] Bertot, J. C., McDermott, P., & Smith, T. (2012). Measurement of open government: Metrics and process. In *System Science (HICSS), 45th Hawaii International Conference on* (pp. 2491-2499). IEEE.

- [9] Gurstein, M. B. (2011). Open data: Empowering the empowered or effective data use for everyone?. *First Monday*, 16(2).
- [10] GovLab. (2015). *Open Data Impact Case Studies*. New York University: The Governance Lab.